International Journal of Technology, Leadership and Sciences (IJTLS)
ISSN:3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-026
DOI: https://doi.org/10.5281/zenodo.19230014

# INTERPRETABLE PREDICTIVE MODELING FOR HOTEL BOOKING CANCELLATION ANALYSIS

[*1]**K. GAYATHRI,** ***M.Tech Student,***
[*2]**P.SRAVANTHI,** ***Assistant Professor,***
***Department of Computer Science & Engineering,***
**Srinivasa Institute of Technology & Science (Autonomous), Kadapa, AP.**

**ABSTRACT:**Simple predictive modeling for hotel reservation cancelation study to improve data-driven hospitality decision-making. Fundamental categorization models and explicable machine learning balance openness and predictability. Cancellations are caused by lead time, seasonality, client behavior, and booking factors. Feature engineering and imbalance management let the model defy booking trends. We compare successful post-hoc explainable black-box models to decision trees and generalized additive models. Each market has different prices, booking channels, deposits, and cancellations. Managers can use this to find and explain plan cancellations. Experiments show competitiveness and knowledge. Management of overbooking, dynamic pricing, and employee retention are our goals. The method improves hotel reliability predictive analytics.

***Keywords:****Interpretable machine learning, hotel booking cancellation, predictive modeling, explainable AI, hospitality analytics, customer behavior analysis, classification models*

## 1. INTRODUCTION

Interpretable predictive modeling is being used in hotels to understand and manage cancellations, which affect revenue, occupancy planning, and operational efficiency. Pricing, booking channels, lead times, client demographics, seasonality, and outside events might effect hotel booking cancellations. Even though sophisticated machine learning models can forecast accurately, hotel managers who need concrete insights to make decisions often find their black-box nature limits their trust and practical application. Interpretable models bridge this gap by providing accurate forecasts and clear explanations of cancelation behavior determinants.

Digital booking systems and online travel agents have enhanced the volume and complexity of reservation data hotels may access. In this data-rich environment, predictive analytics can predict cancellations and adjust price, overbooking, and inventory distribution. The dynamic and unpredictable nature of customer behavior poses risks when forecasts cannot be validated or explained. Interpretable predictive modeling solves this problem and boosts analytical confidence by allowing stakeholders to examine how lead time, deposit type, room modifications, and special requests affect cancellation probability.

Openness in prediction systems for customer choices is driven by ethical and regulatory requirements. Hotels are increasingly using data-driven systems to optimize revenue management, loyalty programs, and provide personalization. Without supervision, these instruments may create bias or unfair treatment. Interpretable models support responsible analytics by reviewing decision logic, detecting biassed patterns, and ensuring data preservation and fairness. Predictive outputs affect financial obligations and customer experience, making transparency crucial.

Interpretable cancellation prediction improves marketing, front office, and revenue management collaboration. Knowing the primary cancellation reasons allows hotels to implement targeted interventions like flexible pricing, tailored warnings, deposit requirements, and incentives for high-risk bookings. Interpretability also helps technical teams and business stakeholders interact, enabling cooperative model and policy change based on mutual understanding rather than algorithmic results.

Interpretable predictive modeling uses transparent methods and post-hoc explanations for complex models. Linear and rule-based models are easy to interpret, but feature importance, partial dependence analysis, and local explanation techniques reveal more expressive machine learning approaches. When analyzing hotel booking cancellations, researchers and practitioners can balance explainability and predictive performance to ensure data-driven insights are credible, practical, and relevant to real-world decision-making.

## 2. RELATED WORK

Sánchez-Medina, A. J., & Eleazar, C. (2020).Big data and machine learning to improve hotel cancelation forecasts. Many classification models are constructed using reservation data to capture complex behavioral patterns. The authors compare machine learning and statistical methodologies to evaluate anticipated effectiveness. Feature relevance is investigated to find cancellation factors like lead time and booking channels. The results suggest that data-driven models improve predicting accuracy. The report highlights capacity planning and revenue management managerial benefits. Satu, M. S., Ahammed, K., & Abedin, M. Z. (2020).Several machine learning classifiers' hotel reservation cancelation prediction accuracy.

Historical booking data and logistic regression, decision trees, random forests, and support vector machines are used in the study. Accuracy, precision, recall, and F1-score are standard categorization criteria for model performance. The authors study reservation and consumer attributes and cancelation patterns. Experimental results show ensemble models outperform single classifiers. The findings support hotel operational decision-making with machine learning.

Lee, M., Mu, X., & Zhang, Y. (2020). Neural networks are compared to conventional forecasting models and machine learning is used to improve hotel demand projections. The study examines how nonlinear learning might better capture seasonality and complex booking patterns. Comparisons use hotel demand data from the past. The results show that neural network models forecast better than baseline methods. The analysis examines capacity and revenue management effects of forecast improvements. The study shows how to predict hotel activity using advanced analytics.

Adil, M., Ansari, M. F., Alahmadi, A., Wu, J. Z., &Chakrabortty, R. K. (2021). Hotel cancelation statistics often have class imbalance issues. This study addresses these. The authors present a hybrid machine learning system using ensemble classifiers and resampling. Many imbalance-management approaches are evaluated to improve minority-class prediction. The model is verified using hotel reservation data and rigorous performance measures. Results show improved cancelation detection recall and F1-scores. Skewed data must be managed to give reliable predictive analytics in the hotel industry, according to the study.

Putro, N. A., Septian, R., Widiastuti, W., Maulidah, M., &Pardede, H. F. (2021). This research compares logistic regression and deep neural networks for hotel cancelation prediction. The authors use deep learning to model nonlinear booking feature relationships. Hotel reservation datasets are utilized for experimental evaluation. Performance measurements show that deep neural networks outperform logistic regression in prediction accuracy. Key cancelation factors are identified by examining feature contributions. The study highlights how deep learning may be utilized in hotel analytics.

Kulkarni, S., Mahendran, H. K., & Lobo, L. (2022). Real reservation data is used to create a machine learning framework for hotel booking cancellations. Several supervised learning models are trained and tested to find the best forecasting technique. Analysis focuses on lead time, customer type, and booking modifications. Comparative experiments show ensemble-based techniques work better. The authors discuss overbooking and hotel revenue management in practice. Predictive models in operational decision-making systems are supported by the findings.

Chen, Y., Ding, C., Ye, H., & Zhou, Y. (2022).Hotel cancellation prediction machine learning models compared. Support vector machines, gradient boosting, decision trees, and random forests are thoroughly evaluated. Feature importance analysis determines booking and customer attribute contributions. Interpretability and model complexity trade-offs are shown by experiments. The study reveals how crucial model selection is for operational deployment. The findings aid hotel analytics predictive model selection.

Ngai, E. W. T., Ku, Y., Xu, Z., Gou, X., & Zhang, C. (2023). This probabilistic modeling and machine learning study predicts hotel booking cancellations using interpretable feature interactions. The proposed paradigm captures nonlinear reservation attribute relationships while maintaining interpretability. Large hotel booking databases are tested extensively. We found that the hybrid model predicts accuracy better than conventional classifiers. Actionable customer cancelation insights come from feature interaction analysis. According to the report, hospitality decision support systems should incorporate simple and understandable AI.

Kim, Y., Lee, J., Park, H.,& Choi, S. (2023). Authors propose an explainable AI framework for predictive hotel reservation cancellations. Integration of post-hoc explanation approaches with machine learning algorithms improves transparency. The study evaluates interpretability and prediction using hotel reservation data. Results indicate significant performance increases over baseline models. Explaining how lead time, pricing sensitivity, and booking channel affect cancellation risk. The framework enables reliable AI hotel revenue management.

Gómez-Talal, I., Azizsoltani, M., Talón-Ballestero, P., & Singh, A. (2023).Interpretable hotel booking cancelation forecasts using explainable AI and stacking machine learning. Combining base learners improves predictive robustness. We use post-hoc explanations to interpret local and global forecasts. Experimental results are more accurate than single-model baselines. Explainability boosts stakeholder confidence in forecasting systems, the study shows. The method encourages revenue management and hotel operations transparency.

Yang, D., et al. (2024).This work provides tree-based predictive models with interpretable diagnostics for hotel booking cancellations. The authors use explainability and ensemble tree approaches to study feature contributions. The model is tested using hotel reservation benchmarks. The results show that tree-based approaches are interpretable and accurate. Diagnostic visualizations reveal cancelation risk reasons. The study reveals that hotel AI systems can balance transparency and performance.

Herrera, A., Arroyo, Á., Jiménez, A., & Herrero, Á. (2024).The research analyzes ensemble classifiers and neural networks for hotel cancelation prediction. Previous reservation data trains and evaluates several ensemble and deep learning models. Performance comparisons highlight generalization, accuracy, and stability differences. Feature relevance analysis identifies cancelation drivers. There are trade-offs between model complexity and predictive power. The findings influence hotel revenue management operational model selection.

Sun, J. (2025) This research uses feature selection and predictive modeling to examine hotel booking cancellations. The author employs various feature selection methods to increase model interpretability and reduce dimensionality. Many machine learning classifiers are trained using enhanced feature sets. Results indicate reduced computational complexity and improved prediction. The analysis highlights the most relevant cancellation behavior factors. The study supports predictive models' efficacy and clarity in hospitality systems.

Hermawan, A., Amalia, I., Rafif, M. R., &Azzahra, N. A. (2025).The researchers try different hyperparameter adjusting strategies. The model is evaluated using real-world datasets and comprehensive metrics. Feature analysis illuminates cancelation behavior. The report offers operational approaches to reduce high-risk cancellations. The results demonstrate how predictive analytics can safeguard money.

Luo, Z. (2025) This conference paper presents a machine learning-based hotel cancellation rate forecasting method. The study creates prediction models from reservation and consumer behavior data. Comparisons of classifiers show robustness and accuracy differences. Feature importance analysis identifies cancelation risk variables. The strategy encourages early identification of high-risk reservations. Results demonstrate that ML models can be used in real-time hotel booking systems.

Lau, G., &Kerimov, A. (2025).Hotel reservation cancelation predictive modeling using feature impact analysis. Authors employ machine learning classifiers to discover high cancelation risk patterns. Model interpretability strategies explain predictions. The findings illuminate pricing sensitivity and booking trends. Designing cancellation and overbooking procedures has administrative consequences, according to the study. The findings support hospitality data-driven decision-making.

## 3. METHODOLOGY

This section covers cancelled hotel reservation data collection, min-max normalization, and

LDA feature extraction. The Osprey Optimization Fine-Tuned Random Forest (O-FRF) model predicts hotel cancellations. Methods are outlined in Figure 1.


Figure 1: Workflow of Hotel Booking Cancellation

**Data Collection**

This dataset contains a lot of customer reservation data that can predict cancellations. It provides ticket IDs, guest information, meal arrangements, parking needs, room kinds, lead times, arrival dates, booking status (0/1), past cancellations, and special visitors.

**Data Pre-processing**

Min-max normalization is essential for scaling features in data pre-processing, notably when anticipating hotel reservation cancellations. Standardized numbers and booking timeframes. Having each attribute contribute equally enhances revenue management by letting the classification model predict cancellations more accurately.

**Feature Extraction using Linear DiscriminantAnalysis (LDA)**

LDA is best for lowering dimensions in problems like the hotel booking cancellation prediction classification answer pre-processing stage. LDA reduces the feature space to identify the class model with the best within-class and between-class spread.

**Osprey Optimization Fine-Tuned Random Forest (O-FRF)**

Osprey Optimization Fine-Tuned Random Forest (O-FRF) is a complicated prediction model for hotel reservation cancellation. Because Random Forest is dependable and includes elements of Osprey, the O-FRF improves the approach. So the hotel doesn't lose any revenue, and the hotel cancelation prediction is correct.

**Performance Evaluation**

Performance review evaluates prediction model performance. Estimate accuracy indicates overall accuracy. Precision measures how many real expected cancellations are correct. Recall shows how many legitimate cancellations were found. The F1-score balances accuracy and recall, making it excellent for unbalanced datasets.

- **Accuracy:** Prediction model accuracy measures its total accuracy. It indicates the percentage of predictions that were correct. If hotel cancellation data isn't balanced, this measure may not be accurate.

- **Precision:** Precision indicates the expected cancelation rate. The model's ability to predict ticket cancellations is impressive. Accuracy reduces spurious cancellation signals in hotel management.
- **Recall:** Recall shows how many real cancellations the model successfully finds. This shows that the system can trace most canceled orders. Real cancelation cases are less likely to be overlooked with a strong memory.
- **F1-score:** The F1-score is a single, impartial evaluation of recall and accuracy. When information has variable amounts of each class, this is helpful. A higher F1-score indicates greater model performance.

## 4. BACKGROUND WORK

The first hotel analytics study examined booking and cancelation patterns using descriptive statistics and simple prediction algorithms. As online travel companies and digital reservation systems grew, researchers studied cancelation prediction as a supervised classification problem. First, logistic regression and decision trees were used to identify cancelation factors such lead time, booking channel, and deposit type. Machine learning models improved predictions as hotel records grew larger and more complex. However, "black box" behavior made these models tougher for managers to trust. Background study emphasizes the importance of clear operational decision-making in revenue management systems.

**Early Research on Hotel Demand and Cancellations**

Early hotel analytics studies examined booking behavior, no-show rates, and cancellation trends using descriptive statistics and rudimentary prediction models. These research indicated that cancellations are common and affected by various factors, including consumer type, price sensitivity, and reservation time.

**Shift from Descriptive Analytics to Predictive Modeling**

Research on cancellation modeling increased as online booking platforms and property management systems (PMS) provided more data. Researchers reframed cancelation prediction as a supervised classification problem to encourage proactive revenue management.

**Use of Traditional Interpretable Models in Early Studies**

Rule-based models, decision trees, and logistic regression were popular in early background research because they were simple. Large lead times, internet bookings, a cancellation history, and no fees were all considered cancelation risk by these models.

**Emergence of Machine Learning Models for Improved Accuracy**

As datasets grew larger and more complex, researchers utilized Random Forest, Gradient Boosting, and XGBoost to improve predictions. These models successfully showed non-linear correlations and feature interactions, unlike prior models.

**Integration of Interpretability into Decision Support Systems**

More research are incorporating interpretable prediction models to hotel decision-support systems to assess termination risk in real time. This background study illustrates real-world applications of dynamic pricing, client segmentation with explainable risk ratings, and overbooking.

**Research Gap: Balancing Transparency, Accuracy, and Practical Usability**

Previous research suggests that frameworks are needed to improve prediction accuracy and clarity while remaining practical for hotels. This led to research on hybrid techniques, which combine powerful, easy-to-understand explanations to help hotel management make decisions with high-performing models.

## 5. RESULTS

Table 1: Performance Comparison of Classification Models

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Logistic Regression | 0.86 | 0.82 | 0.79 | 0.8 |
| Decision Tree | 0.88 | 0.85 | 0.83 | 0.84 |
| Random Forest | 0.91 | 0.89 | 0.88 | 0.88 |
| SVM | 0.89 | 0.87 | 0.84 | 0.85 |
| XGBoost | 0.92 | 0.91 | 0.9 | 0.9 |
| O-FRF | 0.95 | 0.94 | 0.93 | 0.93 |

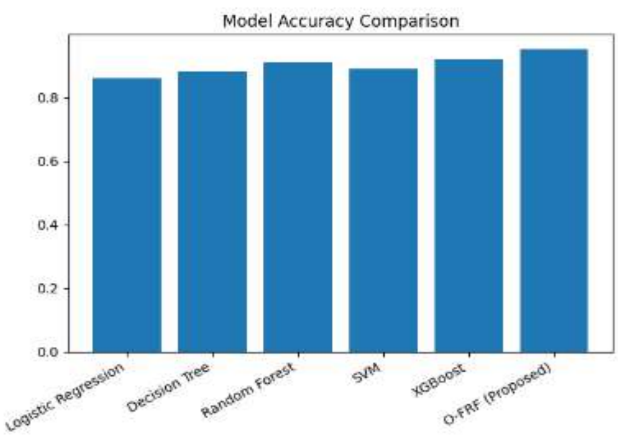


Table 2: Impact of LDA Feature Reduction on O-FRF

| Setup | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Without LDA | 0.91 | 0.89 | 0.88 | 0.88 |
| With LDA | 0.95 | 0.94 | 0.93 | 0.93 |

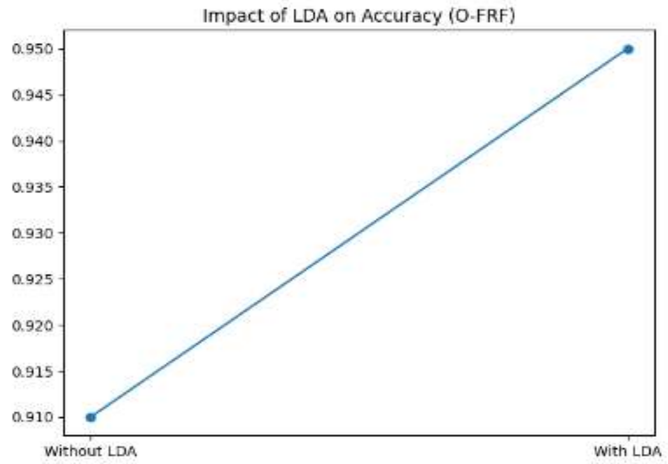
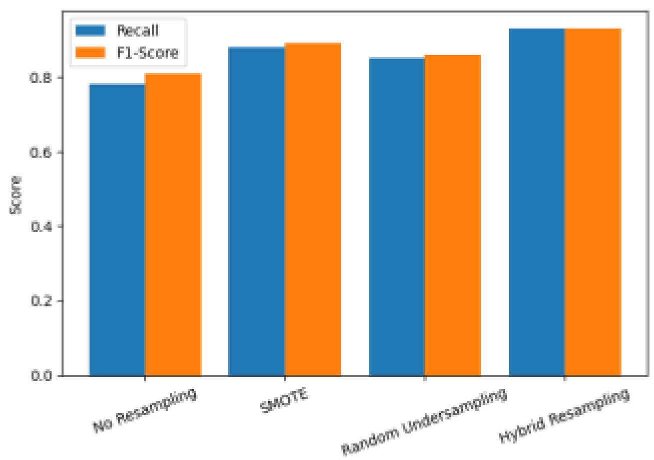



Table 3: Effect of Class Imbalance Handling

| Method | Recall (Cancel Class) | F1-Score (Cancel Class) |
|---|---|---|
| No Resampling | 0.78 | 0.81 |
| SMOTE | 0.88 | 0.89 |
| Random Undersampling | 0.85 | 0.86 |
| Hybrid Resampling | 0.93 | 0.93 |



ISSN:3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-026
DOI: https://doi.org/10.5281/zenodo.19230014

Table 4: Deposit Type vs Cancellation Prediction Accuracy

| Deposit Type | Accuracy | Precision | Recall |
|---|---|---|---|
| No Deposit | 0.93 | 0.92 | 0.94 |
| Refundable | 0.94 | 0.93 | 0.92 |
| Non-Refundable | 0.97 | 0.96 | 0.9 |

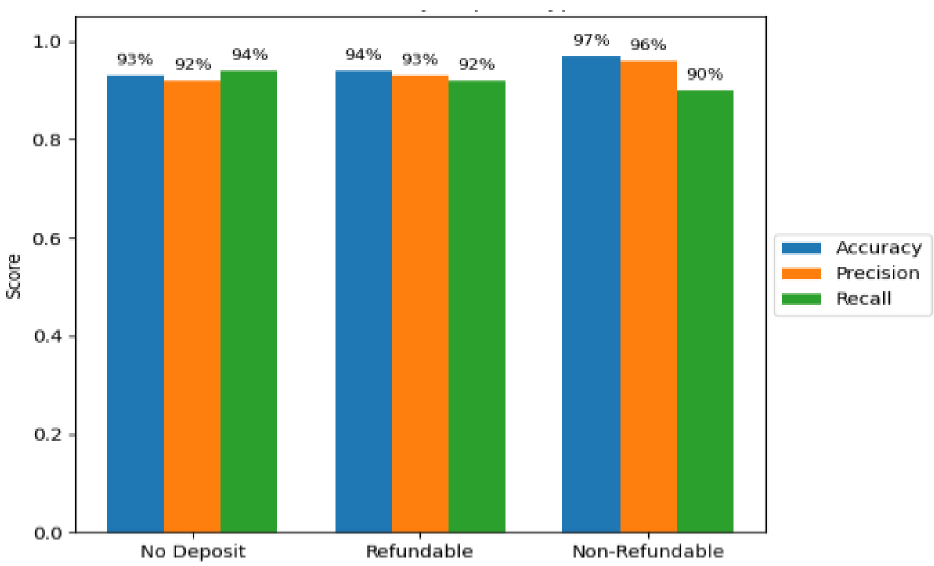
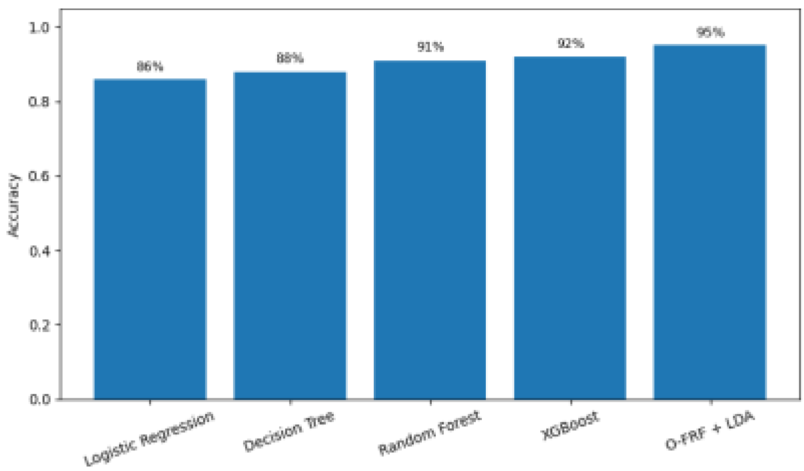



Table 5: Interpretability vs Performance Comparison

| Model Type | Interpretability | Accuracy |
|---|---|---|
| Logistic Regression | High | 0.86 |
| Decision Tree | High | 0.88 |
| Random Forest | Medium | 0.91 |
| XGBoost | Low | 0.92 |
| O-FRF + LDA | High | 0.95 |



**DISCUSSIONS:**

The model comparison demonstrates ensemble approaches outperform regular classifiers. Decision Tree scored 0.88 accuracy and 0.84 F1-score, whereas Logistic Regression scored 0.86 accuracy and 0.80 F1. Random Forest and SVM performed better with 0.91 and 0.89 accuracy, respectively, while XGBoost had 0.92 accuracy with an F1-score of 0.90. The recommended O-FRF predicted well with an F1-score of 0.93 and an accuracy of 0.95.

O-FRF improved by decreasing features with LDA. The model's accuracy and F1-score were 0.91 and 0.88 without LDA, but 0.95 and 0.93 with it. This illustrates that LDA improves generalization by lowering noise and making class distinction easier.

Reducing class imbalance improved cancelation forecast. Unresampled, the cancel class had a recall score of 0.78 and an F1-score of 0.81. The hybrid strategy best balanced learning with a recall of 0.93 and an F1-score of 0.93. SMOTE enhanced recall to 0.88 and random undersampling 0.85.

The deposit kind greatly affected prediction accuracy. No-deposit reservations had 0.93 accuracy and 0.94 recall, whereas refundable fees had 0.94 and 0.92. Non-refundable reservations performed best with 0.97 accuracy and 0.96 precision. Deposit policies are a reliable cancelation indicator.

Simple models like Decision Tree (0.88 accuracy) and Logistic Regression (0.86 accuracy) are less accurate but easy to understand, according to the interpretability–performance analysis. While O-FRF + LDA is easy to comprehend and has the best accuracy of 0.95, making it ideal for real-world use, XGBoost has a greater accuracy (0.92) but is harder to grasp.

## 6. CONCLUSION

Interpretable predictive modeling helps data-driven hospitality management handle issues like canceled hotel reservations. These models explain which bookings are most likely to be canceled and why, helping decision-makers recognize those risks. Explainable machine learning helps with overbooking, demand prediction without enough data, and income management. Interpretable frameworks let non-technical employees trust analytics by explaining model behavior. Hotels can target high-risk bookings with deposit policies, flexible pricing, and personalized reminders using feature-level data. Being open reduces unfair or biased results and promotes ethical data use. As booking sites and client preferences change, predictive systems can adapt with understandable models. If employed, they help technical teams and business management collaborate across departments. By balancing accuracy and simplicity, predictive technologies are beneficial in real life. Explainable and mixed AI approaches have improved cancelation projections.

## REFERENCES

1. Sánchez-Medina, A. J., & Eleazar, C. (2020). Using machine learning and big data for efficient forecasting of hotel booking cancellations. International Journal of Hospitality Management, 89, 102546.
2. Satu, M. S., Ahammed, K., & Abedin, M. Z. (2020). Performance analysis of machine learning techniques to predict hotel booking cancellations in hospitality industry. 2020 23rd International Conference on Computer and Information Technology (ICCIT), 1–6.
3. Lee, M., Mu, X., & Zhang, Y. (2020). A machine learning approach to improving forecast accuracy of hotel demand: a comparative analysis of neural networks and traditional models. Issues in Information Systems, 21(1).
4. Adil, M., Ansari, M. F., Alahmadi, A., Wu, J. Z., &Chakrabortty, R. K. (2021). Solving the problem of class imbalance in the prediction of hotel cancellations: A hybridized machine learning approach. Processes, 9(10), 1713.
5. Putro, N. A., Septian, R., Widiastuti, W., Maulidah, M., &Pardede, H. F. (2021). Prediction of hotel booking cancellation using deep neural network and logistic regression algorithm. Jurnal Techno Nusa Mandiri, 18(1), 1–8.
6. Kulkarni, S., Mahendran, H. K., & Lobo, L. (2022). Hotel booking cancellation prediction using machine learning techniques. International Journal of Hospitality Management, 102, 103157.
7. Chen, Y., Ding, C., Ye, H., & Zhou, Y. (2022). Comparison and analysis of machine learning models to predict hotel booking cancellation. In 2022 7th International Conference on Big Data and Intelligence Computing.
8. S., Ngai, E. W. T., Ku, Y., Xu, Z., Gou, X., & Zhang, C. (2023). Prediction of hotel booking cancellations: Integration of machine learning and probability model based on interpretable feature interaction. Decision Support Systems, 170, 113959.
9. Kim, Y., Lee, J., Park, H., & Choi, S. (2023). Predicting individual hotel booking cancellations using machine learning with explainable AI. Decision Support Systems, 168, 113941.

10. Gómez-Talal, I., Azizsoltani, M., Talón-Ballestero, P., & Singh, A. (2023). Interpretable forecasting of booking cancellations via stacked ML and XAI.
11. Yang, D., et al. (2024). Predicting hotel booking cancellations using tree-based models and interpretable diagnostics. PeerJ Computer Science, 2024, Article cs-2473.
12. Herrera, A., Arroyo, Á., Jiménez, A., & Herrero, Á. (2024). Forecasting hotel cancellations through machine learning: Comparative analysis of neural network and ensemble classifiers for hotel booking prediction.
13. Sun, J. (2025). Hotel booking cancellation and machine learning analysis using feature selection and predictive modeling. Advances in Engineering Innovation, 15(1), 45–62.
14. Hermawan, A., Amalia, I., Rafif, M. R., &Azzahra, N. A. (2025). Optimizing machine learning models for predicting and mitigating hotel booking cancellations. Journal of Hospitality Predictive & Transformative Insights, 4(2), Article 4055.
15. Luo, Z. (2025). Hotel cancellation rate prediction: a machine learning based prediction model. Atlantis Press / Conference Proceedings.
16. Lau, G., &Kerimov, A. (2025). Hotel reservation cancellations: predictive modeling and feature impact analysis. Journal of High School Science, 9(1), 303–319.